2023
ACTIVITY REPORT

Project-Team
AISTROSIGHT

**Viewing neuron-astrocyte pharmacology through digital sciences**

**DOMAIN**

**Digital Health, Biology and Earth**

**THEME**

**Computational Neuroscience and Medicine**

*Inria*

# Contents

# Project-Team AISTROSIGHT

*Creation of the Project-Team: 2023 January 01*

## Keywords

### Computer sciences and digital sciences

A3.1.1. – Modeling, representation

A3.2.2. – Knowledge extraction, cleaning

A3.2.4. – Semantic Web

A3.3.2. – Data mining

A3.4.1. – Supervised learning

A3.4.6. – Neural networks

A3.4.8. – Deep learning

A6.1.1. – Continuous Modeling (PDE, ODE)

A6.1.2. – Stochastic Modeling

A6.1.3. – Discrete Modeling (multi-agent, people centered)

A6.1.4. – Multiscale modeling

A9.2. – Machine learning

A9.8. – Reasoning

A9.10. – Hybrid approaches for AI

### Other research topics and application domains

B1.1.2. – Molecular and cellular biology

B1.1.7. – Bioinformatics

B1.1.8. – Mathematical biology

B1.1.10. – Systems and synthetic biology

B1.2.1. – Understanding and simulation of the brain and the nervous system

B1.2.3. – Computational neurosciences

B2.2.2. – Nervous system and endocrinology

B2.6.1. – Brain imaging

# 1 Team members, visitors, external collaborators

## Research Scientists

- Hugues Berry [Team leader, INRIA, Senior Researcher, HDR]

- Benjamin Vidal [Team leader, Theranexus, Industrial member]

- Audrey Denizot [INRIA, Researcher, from Feb 2023]

- Thomas Guyet [INRIA, Associate Professor Detachement, HDR]

## Faculty Member

- Luc Zimmer [Université Claude Bernard Lyon 1, Hospices Civils de Lyon, Professor, PUPH, HDR]

## PhD Students

- Schayma Ben Marzougui-El Garrai [INRIA, from Oct 2023]

- Lisa Blum Moyse [INRIA, until Sep 2023]

- Andrea Ducos [INRIA, from Nov 2023]

- Florian Dupeuble [INRIA, from Sep 2023]

- Arnaud Hubert [INRIA]

- Nathan Quiblier [INRIA]

- Hana Sebia [INRIA]

## Technical Staff

- Jan-Michael Rye [INRIA, Engineer, Permanent Research Engineer]

- Nicolas Simon [Inria, Engineer, Research engineer CDD]

## Interns and Apprentices

- Mathieu Chambard [ENS RENNES, Intern, from Feb 2023 until Aug 2023]

- Guillaume Girard [INRIA, Intern, from Apr 2023 until Aug 2023]

- Zoe Laffitte [INRIA, Intern, from Apr 2023]

## Administrative Assistant

- Noemie Rodrigues [INRIA, from Nov 2023]

# 2 Overall objectives

## 2.1 Failures in drug development for neurological diseases

The set of available drugs for neurological diseases is both aging and lacking in effectiveness. There remains a very high unmet medical need for treatments in neurology, despite heavy historical investment in the field [25, 56]. A typical drug development process includes a set of successive stages (fig. 1), sequentially evaluating the effects of the candidate drug in vitro on cell culture models, then in vivo on animal models (pre-clinical), after which the mechanistic origins of the candidate drug effect can be assessed in animals (sometimes in humans). The next stage consists of studies in humans based on clinical trials

that are themselves structured in successive phases: phase I to test for safety, phase II to determine the effect of the candidate on a small set of patients and phase III that includes large cohorts of patients and control people in a randomized setting. Each and every stage of this process has significant probability to fail and break off the development of the drug under process. With the recent public health issues (HIV, Covid-19) the general public has mostly been made aware of failures between the successive phases of the clinical trial stage. However, in the field of neurology, the difficulties in developing drug candidates are mainly due to a high failure rate in the clinic: the activity of the drug candidate in in vitro cell cultures or in animal models is very often not confirmed in humans [40, 44].

In recent years, numerical approaches have been proposed in the field, either with mechanistic modeling to predict the response of the cell to the candidate molecule (quantitative systems biology/pharmacology) [51, 57] or with machine learning to identify the impacted (sub)cellular systems or the effects of the candidate drug [65, 46]. However, these approaches are still inefficient to meet the above challenges because they often address a unique scale or modality of interest (e.g., molecular, cellular, preclinical) and lose their predictive power at other scales (e.g. clinical, i.e. the patient). The main methodological objective of AIstroSight is to develop quantitative systems biology and Artificial Intelligence (AI) approaches able to embrace several of these scales.



Figure 1: A schematic view of a drug development process
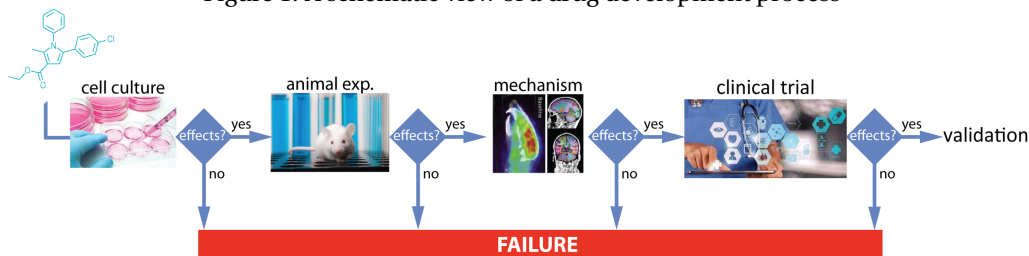
## 2.2   Main deliverables

Our overall goal is to develop innovative numerical methods for neuropharmacology that will provide us with levers to accelerate and derisk the early stages of drug design. As a main deliverable and proof of concept of the efficiency of these methods, our ambition for the first four years of the project is to identify a handful (2 to 4) of new candidate drugs against neurological diseases.

## 2.3   Overview of the AIstroSight roadmap

To improve the probability of success of drug candidates in neurology, we integrate complementary information offered by data harvested at different spatio-temporal scales (fig. 2): from the inside of the cell (molecular and cellular biology) to the whole brain (imaging) and even to a population of patients (hospital data), using numerical tools coupling mechanistic models with dedicated AI approaches. In a way, our strategy is to break down the classical stratification silo of Fig. 1, in which literature search, in vitro cell culture, in vivo preclinic studies and in vivo clinic studies are viewed as a sequential multi-stage process. Instead, we propose an integrated machine learning framework into which all those data are combined to predict the effect of a candidate drug molecule.

AIstroSight develops innovative numerical approaches to integrate these information sources into a coherent stream of data and expert knowledge, combining the analysis of experimental observations with reasoning (of different kinds). Currently, these tasks are carried out in isolation and their reconciliation is devolved to biologists/physicians. The originality of the AIstroSight contributions are approaches that automatically carry out this reconciliation to assist biologists/physicians.

Since AI algorithms are often black-box tools, we also develop mechanistic modeling approaches (multiscale quantitative systems biology/pharmacology) to produce explanations for the predictions of the AI algorithms, that can be rooted in neurobiology. Another important aspect of AIstroSight is to widen the focus of neuropharmacology beyond neurons, that constitute only one half of the nerve cells in the

Figure 2: A schematic for the scientific roadmap of AIstroSight



multiscale - multimodal data integration    multiscale mecanistic modelling

$$(2\pi a r_i)^{-1} V_{xx} = C_m V_t + G_{Na}(V)(V - E_{Na})$$
$$+ G_K(V)(V - E_K) + (V - E_L)G_L$$
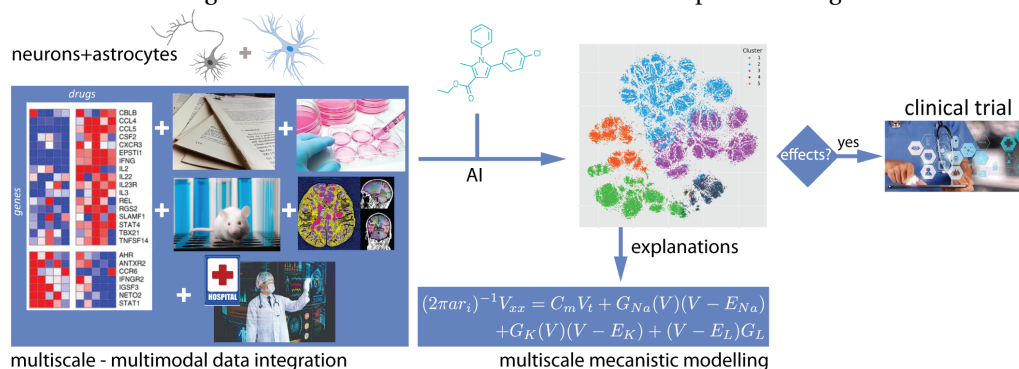
brain, and also take into account the other half, that is made up by glial cells and their interactions with neurons. In particular, we consider the pharmacology of astrocytes [67], one major subtype of glial cells, in interaction with the pharmacology of neurons.

## 2.4  Principles of the AIstroSight partnership

To accelerate cross-fertilization between digital science and medical research, AIstroSight will be located on the East Hospital Campus of the Lyon University Hospital, the "Hospices Civils de Lyon" (HCL), from 2024. We will also benefit from our strong association with CERMEP, the preclinical and clinical in vivo imaging platform of the HCL. In 2024, the whole team is indeed expected to move to Lyon's neurology hospital, located just across tA joint team with the HCLhe street of CERMEP.
CERMEP is also affiliated with University Claude Bernard Lyon 1, Inserm and CNRS. This provides us with an exceptional environment for the engineering of brain biochemical imaging methods that allow the study of the effect of molecules on the whole brain (fMRI, PET, fUS) and the analysis methodologies for these images. The CERMEP also hosts team BIORAN of the "Centre de Recherche en Neurosciences de Lyon" (CRNL) laboratory, that has expertise ranging from the chemistry of candidate-molecules to their biochemical assays, from radiolabelling to animal PET/MRI imaging and from preclinical models to first-in-man studies in patients. The modeling expertise on the binding between candidate molecules and receptors (structural biology, docking) is also present at the CERMEP.
As a joint team with the HCL, part of AIstroSight technology developments is intended to be integrated into the hospital information system developed during the last decade by the HCL for patient management. This is in particular the case of the development of "Multi-patient query for care pathway characterization and clinical trials". Beyond participating in the HCL's mission as an innovation leader in digital health, AIstroSight also represents an opportunity for the HCL to reinforce its infrastructure for the organization of clinical trials, for instance in cooperation with pharma/biotech companies like Theranexus. Like other teams of Inria Lyon, AIstroSight is intensively implicated in the "AI innovation department" ("Pôle de Développement IA") that Inria Lyon and HCL are supporting.
Finally, to ensure the impact of our works on pharmacology and provide it with potential industrial exit routes, the AIstroSight partnership also includes an industrial partner, Theranexus, a French biopharmaceutical company (an SME) that develops drugs for the treatment of nervous system diseases with an original focus on both neurons and astrocytes. Theranexus is listed on the Euronext Growth market in Paris and its social headquarters are located in Lyon. A biotech company, the expertise of its members is entirely in the experimental aspects, not in the digital. Theranexus brings to AIstroSight experimental data (experimental cell biology and brain imaging for pharmacology), and provides their expertise for the development of the digital tools needed to analyze these data. In return, the objective is that the output of these digital tools reveal novel drug targets or novel candidate molecules that Theranexus may decide to use to develop new treatments, starting with the necessary clinical trials. Importantly, the fact that these candidate drugs have been selected from an innovative numerical approach strongly consolidates the credibility of their development on the pharmaceutics market. In addition, Theranexus brings to AIstroSight its know-how and industrial expertise on the development of drug candidates up to

the market and its strategic knowledge of the neuropharmacology industry. In this operating scheme, Theranexus is therefore the preferred partner for the early-phase transfer of the molecules that AIstroSight could identify.

Independenlty from AIstroSight, Theranexus and BIORAN have a longstanding collaboration together, in particular in the framework of an ANR- and AURA Region-funded joint laboratory (LabCom) called « NeuroImaging for Drug Discovery (NI2D)» that aims at the development of gliopharmacology using preclinical imaging tools (PET/MRI/brain ultrasound). This LabCom is also hosted within the CERMEP premises. NI2D aim at developing preclinical neuroimaging techniques (in animals, mainly fMRI, PET and fUS = Functional Ultra-Sound) for the evaluation of drug-candidates. AIstroSight develops numerical methods capable of integrating data at multiple scales for pharmacology, data that include imaging but also molecular data (intracellular signaling, omics data) or clinical data (biology, treatments, medico-administrative). We thus benefits from the imaging data and methodologies of NI2D. The two are therefore complementary, especially since we both share a strong interest in neuron/astrocyte interactions.

## 3    Research program

### 3.1    Characterizing the mechanisms of action of candidate drugs

Drug screening, either in vitro or in silico, generally does not provide an explanation of the mechanism by which the identified drug acts at the cellular level. However, this information is crucial (e.g., with respect to patients or health agencies), and the algorithms used for the screening must be made explicable. Our goal here is to use mathematical models and their hybridization with machine learning to provide explanations on the mechanisms of action of a candidate molecule.

We develop mechanistic models of regulatory networks or intracellular signaling pathways specific to the action of the candidate drugs identified by the screening. Those models predict the spatio-temporal evolution of the concentrations of the molecular species involved in the modelled pathways using classical reaction terms from biochemical kinetics and mass-action laws (first order reactions, bi-molecular reactions, Michaelis-Menten, Hill kinetics...). Depending on the importance of intracellular spatial gradients and biochemical noise, space and stochasticity is accounted for, thus resulting in models based on reaction-diffusion equations, stochastic or ordinary differential equations, or other related formalisms (Gillespie algorithm, flow analysis). These models allow us to simulate in time and/or space of the cell the mechanisms that govern the dynamics of the implicated molecules and how this dynamic is altered by the selected drug. The aim is to use such mechanistic modeling to produce explanations for the predictions made by the statistical learning techniques that are used in the other sections. It is unlikely that these mechanistic models in themselves allow us to decipher the totality of the molecular mechanisms involved, but they provide critical information to properly adjust the laboratory and clinical experiments.

To be efficient, this approach demands that we maintain an effective knowledge basis and expertise on the fundamental molecular mechanisms at play at these spatial scales and their modelling. To build and maintain this expertise, we rely on existing long-term collaborations between AIstroSight members and experimental neuroscientists -electrophysiologists- or neuropharmacologists on the intracellular signaling networks at play in neuron function or neuro-astrocyte interactions.

- *Agonist bias in GPCR*: G protein-coupled receptors (GPCR) are currently the largest family of molecular targets for potential new drugs [22]. GPCR are cell-membrane receptors ubiquitously found in all mammalian cells, but in particular in brain cells (neurons and astrocytes), where they control a large repertoire of neuronal and astrocytic responses to a variety of external stimuli and molecules. Bias antagonism refers to the observation that two ligands of the same GPCR can activate very different cell responses [47]. This phenomenon is still not understood, but it is one possible path towards the development of new drug discovery and has been already proposed and stated to be explored, in particular by members of AIstroSight (B. Vidal, L. Zimmer) [68, 55]. Our objective here is to build realistic mechanistic models of GPCR-based cell signaling in the neuronal intracellular space. We plan then to use this/these models to propose molecular mechanisms to explain the experimentally observed biases. A first idea to explore is the hypothesis of a local subcellular compartmentalization of the signaling molecules over and close to the cell membrane (so-called nanodomains). Experimental validation of the main model predictions is then to be performed using brain imaging modalities available at CERMEP (TEP, MRI, fUS).

- *Synaptic plasticity*: Synaptic plasticity, the long-term adaptation of the efficacy of a synapse according to the activity of the neurons and astrocytes composing this synapse, is thought to underly learning and memory at the cellular scale [64]. We have been enjoying a very fruitful collaboration with Laurent Venance's lab (INSERM U1050, CIRB, Collège de France, Paris) on the subcellular mechanisms at play in learning and memory formation by synaptic plasticity [32, 34, 35, 33, 41, 70, 45]. Current work focuses on the control of synaptic plasticity mechanisms by endocannabinoids and its implication in fast learning and on metabolic regulation of synaptic plasticity by astrocytes. This collaboration is funded by ongoing ANR project EngFlea (see below).

- *Calcium signaling in astrocytes*: Calcium signaling in the terminal branchlets of astrocytes is thought to be crucial for astrocytic functions and neuron-astrocyte interactions [24]. We are studying the local dynamics of calcium signaling in terminal branchlets of astrocytes and their interaction with synaptic activity in collaboration with U. Valentin Nägerl's lab (CNRS UMR 5297, Bordeaux) for experimental (subcellular) data with supra-resolution microscopy [36]. Recently, a collaboration with Erik de Schutter's lab (Okinawa Institute of Science and Technology, Japan) has also been set up to develop new efficient modelling tools (stochastic reaction-diffusion systems) in realistic 3D geometric meshes based on the simulation framework they develop, STEPS [23, 37].

- *Multiscale modelling of the effects of a candidate drug on neurons and astrocytes*: To model the cellular effect of a candidate drug, the main molecular systems impacted by the drug, and thus to be accounted for in the mechanistic model, are isolated from the cellular signature data and literature exploration. Imaging data, by specifying the brain areas and structures mainly targeted by the candidate drug, helps refine these models using specific parameters. Whereas in the first models, the observation (and modelling scale) corresponds to a subcellular domain (one synapse, +/- a dendrite or the main astrocytic process in the neighborhood), we search to progressively scale up those mechanistic models from the intracellular scale of a single cell to the scale of a population of interacting brain cells, neurons and astrocytes. To do so, we explore model simplification /reduction methods, including those combining machine learning and dynamical systems modelling (see below). In the long run, this large-scale mathematical model will produce a digital twin of the pathology that will allow us to explain why the candidate drug has a positive effect on the disease. Calibration is based on fUS and fMRI imaging data in rodents obtained in the framework of the NI2D LabCom. This data provides us with quantitative measurements of the effects of microscopic perturbations by pharmacological agents or by external stimuli (e.g., visual) on the variation, correlation and spreading of cortical activity over the whole brain.

- *Astrocyte roles on brain imaging signals*: Although it is now widely accepted that astrocytes play a role in brain processes and pathologies, the exact perimeters of their roles remain to be delimited. For instance, variations of the signals measured by brain imaging methods (fMRI, PET, fUS) are still largely interpreted as variations of neuronal activity. Available experimental data however indicate that astrocytes also impact those signals but it not clear yet how they do it. A precise and quantitative answer to this question would allow us to use brain imaging to monitor not only the local activity of the neurons but also of the astrocytes. Such a feature would be precious in our framework of astrocyte pharmacology but it demands the development of new mathematical models. Existing models of fMRI signals, for instance, are either too crude to incorporate a separate astrocyte action (balloon models [43]) or are limited to the role of astrocytes as energy suppliers of the neurons (astrocyte-neuron lactate shuttle [49]). Our objective here is to start from a microscopic and mechanistic model of neuron-astrocyte-blood vessel interactions and use multi-scale modelling methodologies to obtain a large-scale model of astrocyte-neuron impact on a subset of brain imaging technics (fMRI, fUS), with explicit parametrization of local neuronal and astrocyte activities. Here again these models are calibrated using fUS and fMRI imaging data in rodents, in particular using pharmacological agents that are known to specifically silence the astrocytic population or a neuronal population in a given brain area. A crucial step is the development of a detailed, microscopic model of the astrocyte endfeet, the specialized astrocyte processes that respond to and control local vascular diameters. This model will provides us with causal mechanisms able to interlink neuronal electrical activity, astrocytic calcium activity and local blood flow. It is to be seen as a first stage towards understanding the implication of astrocytes in variations of neuroimaging signals.

**Methodological challenges:** The biological systems to be considered to explain drug effects on pathologies are not only very complex but also only partly understood by neurobiologists themselves. Therefore, the available biological knowledge on these systems is constantly evolving. Since we cannot know in advance what systems are affected by the candidate drug, a major difficulty for modelers is preparedness,

i.e. maintain a level of expertise on the biology and modelling state-of-the-art of a wide range of those systems. This is the reason why the first three projects above are crucial to the success of our proposal. The most important challenge we face is that of multiscale: causal data are mainly molecular but many observations are macroscopic (e.g. brain imaging). Traditionally, linking these two scales requires the development of new theories (e.g., homogenization, population boundaries, etc.), a slow and rather hazardous process. The availability of more and more important computing resources allows to consider "brute force" approaches in which all scales of time and space are numerically simulated (cf Blue Brain Project). But the results are often as difficult to interpret as the animal experiments that these simulations emulate. Instead we consider recent advances in hybrid digital-AI systems (physics-informed neural networks [60]), in particular equation discovery methodologies [59]. These methods usually use sparse regression techniques to select in a library of nonlinear terms and operators, those that, when composed, provide the best description of the data [28]. Our idea is to generate a large number of numerical simulations at the microscopic scale of the kinetics of the biochemical reactions concerned, for example by the spatial Gillespie algorithm, and then to aggregate them at a higher spatio-temporal scale using, for example, averages at a space and time grain much higher than the spatio-temporal resolution of the initial microscopic simulations. The idea is then to use equation-discovery algorithms to infer a set of differential equations (and associated parameters) capable of describing these higher scale space-time kinetics. The resulting reduced model is then replicated in each cell of the cell population model. If successful, this model reduction process can even be reiterated at the upper scale to simulate the effect of the molecule on large brain areas. Of course, the risky and difficult nature of this objective makes it a long-term goal. If need be, alternative meta-modelling technics will also be considered when applicable (RKHS, model-order reduction).

## 3.2 Multi-patient query for care pathway characterization and clinical trials

Real-world data, especially the data that is routinely collected by hospitals (medical reports, hospital records...), provides rich information about possible links between patient information (demographic, pathological, life style), drug exposures and health events. In the context of drug development, this data can be useful at three stages. During the search for a new drug, they can be used to enrich cell culture data or imaging data. In this case, one can query patients that have been treated for the pathology in question and integrate their clinical data in the in-silico screening. This approach is presented below in the framework of data integration.

**Electronic Health Records query algorithms:**

Efficient patient query can also be used at the very initial stage of drug discovery: the assessment of the feasibility of drug development projects. Indeed, part of the pathologies we target are rare diseases. In this context, one has to make sure at the very early stages that the pathology in question is not so rare that the number of patients is too low to allow clinical trials, or that its description in terms of physiopathology is mature enough for the clinician to be able to diagnose it with good probability. We thus develop patient query algorithms on clinical data from hospitals (electronic health records, EHR), in particular of the HCL, that allow us to characterize the care pathways of the patients before and after diagnosis. They provides us with answers to many questions related to the clinical picture of the pathology, its genetic underpinnings, its prevalence rate, the typical care pathway of a patient with this pathology, the delay of diagnostic, the frequency of diagnostic errors etc. Answers to these questions are crucial to determine early on whether the drug discovery project is feasible. We aim at developing query algorithms and software pipelines for EHR that can provide us with tools able to answer these questions efficiently.

Efficient EHR query algorithms are also very useful at the final stage of the clinical trial itself (Fig. 2), where they can be used to finely select what patients should be integrated in the trial. Indeed, a major change of paradigm in medicine in recent years is the acceptation that the response of a group of patients to a drug treatment exhibits strong variability. The source of this variability is diverse [69]. The definition of the pathology itself as a unique coherent class can be misleading and actually incorporate a range of different sub-classes of pathologies/disorders. The response to a drug also depends on how the patient's body affects the drug before it reaches its target organ/cells (pharmacokinetics). At the cellular level, the response can also vary because of inter-individual differences in gene sequence and receptor/protein structure (pharmacodynamics). Therefore, individual drug responses depend on the patient genes (pharmacogenetics) but also on more social factors (age, sex, anterior medical record,

lifestyle, habits, exposure to pollution…). In any case, the strength of this variability is believed to be a major cause of failure for clinical trials, in particular in neurology and psychiatry [56, 50]. The goal of "stratified medicine" in this perspective is to subdivide the available group of patients into a number of subgroups so that the response of each subgroup is less variable than the whole [38]. Our objective is to develop computational tools and software packages able to stratify hospital data to assist in the selection of patients to be included in an evaluation protocol for a clinical trial or the building of a research cohort.

**Computational phenotyping:**
The task of querying patients according to a predefined criterion from a large population of EHRs is sometimes referred to as "computational phenotyping" [61]. It remains a time-consuming and challenging task with complex criteria because the query is to be addressed within multiple document types and across multiple data points, in EHRs that usually comprise both structured and unstructured data. The computational challenges raised by patient query with complex criteria are therefore considerable (integration, query, analysis, privacy). Software tools (i2b2, ACE [29]) have been proposed to query patients for cohorts or clinical trials based on EHRs but they can hardly be used by most of the physicians because they require advanced knowledge of the data in computer science terms (format, encoding). Moreover, our objective is to provide clinicians with tools able to manipulate these complex data together with medical concepts (e.g., exposure to a drug, treatment, or occurrence of a pathology). Data abstraction capabilities must therefore be integrated to automatically enrich the data using phenotype libraries that can be intuitively mobilized by the clinician. In analogy with bioinformatics workflows, we create workflows for computational phenotyping.

In cases where we already know how to stratify, the issue is not a learning problem but rather a query problem. On the other hand, when this is not the case, we have to develop methods to build these homogeneous subgroups, and in this case it is a question of (unsupervised) learning: the training criterion becomes a measure of cluster homogeneity. Two competing approaches can be thought of in order to create the building blocks of the workflow: 1) machine learning approaches that allow the construction of abstract patient phenotypes from massive data; 2) approaches inspired by both timed systems modeling and knowledge reasoning that rely on formal descriptions of computational phenotypes to enrich the data. The interest of formal descriptions is to be able to represent the whole data transformation in a formal way. This abstract representation of the construction of a cohort facilitates its understanding by users and its reproducibility (FAIR principle). On the other hand, they allow again to exploit intimately the formalized knowledge of the domain, but they also become objects that can be manipulated by reasoning tools. The use of semantic web technologies can therefore be an interesting tool for representing data, knowledge and their processing in order to propose query tools that guide the clinician through the knowledge.

**Methodological challenge:**
The challenge is to make these formal descriptions highly expressive and to ensure efficient processing of massive data. On the long run, we plan to take inspiration from the approach called "Ontology-Mediated Query Answering" which consists in using ontologies to mediate the query of a database by ontologies [27]. In this context, a computational phenotype is seen as a query. The difficulties encountered with observational data is the semantic gap between the available data and the medical concepts that are interesting to manipulate. This gap may be bridged by automatic reasoning that exploits expert knowledge to relate different abstraction levels.

Since computational phenotypes are difficult to formalize, the challenge is to support clinicians in defining them. In other words, the challenge becomes to abstract phenotypes from clinical data. We plan to combine automatic reasoning methods and data analysis. The first research direction we propose is the exploration of a symbolic approach parallel to the work by Tijl de Bie [26] or by Silberschatz [63] on the notion of "subjective measure of interestingness". This approach was developed to identify user-relevant statistical analysis results by means of a statistical model to evaluate the novelty of the extracted patterns (a priori knowledge model). Symbolic approaches can be combined in a similar way by using symbolic data analysis methods such as pattern mining, and by relying on formal models of the system as a priori knowledge. Patterns that are not "explainable" by the formal model are potentially new or of particular interest to the user and will thus be extracted. This approach offers an original entry point to deeply integrate knowledge-based reasoning into pattern extraction methods. The research challenge here lies in combining formalized knowledge with experimental data. It may be implemented using the declarative pattern mining paradigm, that uses solvers to address the pattern mining task. The proofs of concept

on the notion of novelty will open the way to more complex reasoning such as planning that can be used to integrate complex behaviors in biological systems, such as interaction networks. The second research direction we propose is based on recent machine learning techniques. Unsupervised ML has been applied to patient phenotyping, i.e. the discovery of phenotypes from EHR data, including temporal phenotyping [71]. Our objective is to combine such kinds of algorithm with semantic knowledge to guide the discovery toward meaningful computational phenotypes. Indeed data embedding techniques can integrate ontologies to enhance data semantics [42].

On the long run, the methods developed above may be reunified to address the problem of drug discovery at both the biological and the body scale. This justifies the coherence of the methodological approaches (Semantic Web and machine learning) that are developed in the two objectives.

# 4 Application domains

## 4.1 Targeted Pathologies

The list of pathologies that are of interest for Theranexus in the framework of AIstroSight is given in the stand-alone "convention d'équipe-projet commune" of AIstroSight. It comprises roughly 30 rare diseases of the central nervous system, including lysosomal pathologies, neurological genetic diseases, rare diseases due to $\alpha$-synuclein accumulation or rare demyelinating pathologies. This list may be updated, subject to the prior joint written consent of AIstroSight partners. For the pathologies in this list, a specific regimen is defined in terms of IP and legal affairs. AIstroSight members are allowed to work on pathologies outside this list without any restriction but with a different legal regimen vis-à-vis Theranexus. In agreement with the company, we have selected two pathologies from this list as priority objectives, on which we will start our work: Rett syndrome and Niemann-Pick Type C disease. Both pathologies are neurodevelopmental diseases caused by mutations in a single gene: mutations in methyl-CpG-binding protein 2 (MeCP2) for Rett [30] and in NPC for Niemann-Pick C [58]. MeCP2 mutation in Rett syndrome causes slowed brain growth, a progressive loss of movement, motor control abilities and language in the children and can also cause heavy breathing problems, epileptic-like seizures or intellectual disabilities, among others [48]. NPC mutations in Niemann-Pick type C causes accumulation of cholesterol and other fatty acids inside lysosomes, including in brain cells. The symptoms are highly variable, ranging from defects of developmental and motor progression, difficulties in learning, speech or swallowing, to cognitive impairment or psychiatric symptoms [66]. Restoring NPC expression in astrocytes significantly increases survival of mice models of the disease, suggesting that astrocyte dysfunction is involved in disease progression [72]. The two diseases also have in common that they are rare neurological diseases of children (frequency $1/10^4$ and $1/10^5$, respectively) and that there is no known effective treatment. This rarity has strong consequences on the numerical tools that can be used to find potential pharmacological targets.

## 4.2 In vitro and in vivo experimental models of neurology diseases

Many of the diseases that are of interest for AIstroSight are rare diseases. This means that the volume of experimental data and the basic understanding of the pathology at the (sub-)cellular level may be too limited for the machine learning or mechanistic modeling tools that we plan to use. For example, it is known that the NPC mutation in Niemann-Pick type C induces morbid cholesterol accumulation in cells but the molecular function of NPC in cholesterol metabolism is not clearly understood [58]. Similarly, MeCP2, the gene mutated in Rett syndrome, is an epigenetic regulatory factor (DNA methylation) whose mutation theoretically impacts the expression of a large number of genes but it is not clear which ones are most involved in the symptoms of the disease [48]. Although molecular (omic) studies have been published for both diseases [31, 62], their molecular contexts are still unclear.

Our goal here is to generate additional preclinical molecular and imaging data to better delineate the perturbations that these diseases cause at a cellular and tissue level. We introduce into cultured cells the same deficits as those observed in patients. Transcriptomic analysis of the effect of this manipulation gives us information on the implicated molecular networks and its major molecular consequences. In parallel, we induce these same perturbations in vivo in rodents. Observing these animals using brain imaging techniques (fMRI and fUS, possibly PET) gives us a more macroscopic view of the effect of

the mutation (affected brain areas, nature and amplitude of the modifications, change in response to treatments or stimuli etc, see below).

**Methodological challenges**: Developing experimental models of pathologies can be a very difficult task for pathologies that are due to the conjunction of multiple factors, when the molecular alterations at the origin of the pathologies have effects over a very large range of cellular processes or when comparison of the phenotype of the experimental model with its human counterpart is ill-defined (psychiatric diseases, for instance). To mitigate this risk, we develop experimental models only for pathologies that are well-defined in molecular terms, like for Rett or Niemann-Pick type C for a start. We use viral vector strategies (mostly shRNA-mediated gene silencing or possibly CRISPR-based gene editing via adeno-associated viruses, AAV) to manipulate the sequence or expression of the target gene. We start with cell lines that are easy to grow and analyze using omics approaches, and then use neurons and astrocytes differentiated from human pluripotent stem cells. This approach is also used in vivo by locally injecting the viral vector into a given brain region of an animal model, to genetically modify a particular cell type by using a specific promoter. We should therefore be able to control the area of the brain in which the genetic manipulation will be induced (e.g. visual cortex or cerebellum) as well as the type of cells targeted (neurons vs. astrocytes, for example). Of course, like all experimental models, each model taken separately has its limitations: the genes expressed by cells in culture are not necessarily those expressed by these same cells in vivo, the effects of gene silencing in a rodent are not necessarily transposable to humans, etc. However our hypothesis is that by combining these different modalities and scales of data (see above), it should be possible to better predict the effect of a potential treatment. The molecular and cellular biology technologies to be mobilized here (in vitro and in vivo mutagenesis, cell culture, proteomics) are tools routinely used by Theranexus. The expertise on the use of medical imaging to observe the effects at the brain level is provided by CERMEP and benefits from the advances of the NI2D LabCom.

## 4.3 Identification of multi-source multi-scale biomarkers

Recently, Theranexus changed its pharmacological strategy, from a strategy mainly based on the re-positioning of pre-existing drugs to a technology based on antisense oligonucleotide drugs. These technologies rely on the ability to design on demand short RNA sequences that specifically bind the mRNA of a gene target, and knock it down after recognition by the RNase H1 enzymes present in all cells, or modulate its translation or splicing via steric hindrance . Pharmacological intervention thus consists in searching for a gene target able to correct the molecular perturbation caused by the disease and to synthesize an antisense oligonucleotide able to specifically bind this target gene. Note that the technology currently in clinical use does not (yet) provide ways to specifically target a cell type or a brain region.

Our first objective is to develop digital tools to model the molecular networks perturbated by the pathology of interest, and use this model to identify a gene or protein in the network the modulation of which would correct the perturbation caused by the pathology. These models are based on molecular data, in particular transcriptomics and metabolomics data. The set of data includes data derived from cell cultures as described above, that we augment with molecular data from the literature related to the pathology or more generic public, open access databases of transcriptomic responses to perturbating molecules, like CMap or the LINCS L1000 data repository. The latter, for instance currently includes the effect of close to 40,000 small perturbating molecules on 12,000+ genes of more than 200 cell types. We aggregate these data and use them to infer the gene interaction network, the metabolic network and/or the signaling network impacted by the pathology. Metabolic networks are important for instance for Niemann-Pick type C, to conciliate perturbations of the lipid metabolism with those of the gene expression network. This provides us with a view of the pathology at the molecular scale.

**Integration of neuroimaging data:**

A major objective of AIstroSight is to augment these molecular data with medical data, in particular brain imaging data and hospital data. We complement molecular data with data coming from the analysis of brain imaging (fMRI, PET, functional ultrasound brain imaging) i.e. with functional networks between brain areas targeted by the molecule or quantitative measures of radioligand binding. Most of this imaging is done in rodents (preclinic, see above) but a subset of human imaging data is also used. These imaging data are obtained by our collaborators from the CERMEP platform.

These different neuroimaging methods provide meaningful and complementary information for understanding the functional or molecular effects of drugs in the brain:

- Positron emission tomography (PET) enables to visualize and measure the concentration of a specific radiotracer, with nanomolar to picomolar sensitivity. Numerous brain PET radiotracers have been developed over the years, enabling to study various molecular processes (such as the synthesis and release of endogenous neurotransmitters, the density of receptors, transporters or proteins aggregates, neuroinflammation and cerebral metabolism) both in animals and humans in a non-invasive way. We especially focus on the measurement of cerebral glucose consumption using [18F]FDG, a radiolabeled glucose analog, while also taking advantage of the rich information that can be obtained using other brain radiotracers when needed. Indeed, the asset of [18F]FDG-PET imaging is to be relevant for virtually all drugs that are expected to be active in the brain, since the cerebral glucose uptake is related to neuron and astrocyte activities. In addition, this neuroimaging technique can be used in awake freely-moving animals, getting rid of the anesthesia or stress confound in preclinical studies.

- Functional magnetic resonance imaging (fMRI) enables to follow the dynamic changes in the BOLD signal (for "blood-oxygen level dependent"), which is also related to the brain activity, in a non-invasive way. It has a high temporal resolution (2 to 3 seconds) as compared to PET imaging (several minutes) and therefore can be used either to measure the time series of BOLD signal changes after injection of a drug (called "pharmaco-MRI") or the changes in functional connectivity occurring after neuropharmacological stimulation. Functional connectivity is defined as the correlation in activity over time between different brain areas; this concept has largely become prominent in neurosciences over the years, and it is known to be modified in many physiological or pathological states. fMRI can be used in animals and humans, similarly to PET, but usually requires anesthesia for preclinical applications.

- Functional ultrasound imaging (fUS) is a much more recent imaging technology. It provides access to the dynamic measurement of cerebral blood volumes changes, which are more straightforwardly related to the brain activity as compared to the BOLD signal. Moreover, its spatial, temporal resolution and sensitivity are unmatched by the previous techniques, and it can be applied to freely-moving awake animals in real-time, in correlation with a particular behavior. However, it is currently limited to imaging in 2-dimensions (one brain plane at a time) and is mostly suitable for animals. Therefore, fUS imaging is a complementary way to study brain activity in small animals in the context of preclinical neuropharmacology.

**Integration of clinical data:**
We also plant ot integrate hospital data from the Hospices Civils de Lyon according to availability and pathologies. Hospital data provide access to rich information on possible links between patient information (demographic, pathological), drug exposures, health events or biological sample analysis (e.g., blood markers). Our goal is to integrate brain imaging and hospital data with cellular signatures to enrich them with information at the individual scale in a form that can be analyzed with machine learning (clustering, classification) or data mining (pattern matching) methods.

**Methodological challenges:**
A first challenge resides in the nature of action of antisense oligonucleotides, that often work by knock down/loss of function. It is not straightforward to design such a strategy in the case of a pathology that is due to a mutation that already suppressed the effect of a gene. That is precisely where numerical models of the involved gene expression and metabolic networks are important because they can be systematically assessed for the effect of gene suppression, thus providing a quick in silico screening of the potential targets. However, part of this program implies typical bioinformatics processing steps: analysis of transcriptomic networks, network reconstruction, conciliation between transcriptomic and metabolic networks… We currently do not have this expertise in the team. Therefore we leverage collaborations with local experts of the field to get the necessary operational knowledge, including experts of brain transcriptomics analysis (MeLiS lab in Lyon).

Another difficulty lies in the heterogeneity of these multiscale data, their highly categorical character, the large dimension of the corresponding variable space and often, the small number of observations. Moreover, cellular signature data are intrinsically very noisy and can have low reproducibility [52], a caveat that feature selection may improve, at least in part [39]. Class imbalance can also be strong. Finally, each type of available observation (molecular networks, imaging, hospital) gives a partial, fragmented and incomplete view of an abstract complex biological system. This is a partial view because each type of observation provides data at a given spatio-temporal scale, for a certain locus. This is a fragmented view because the data will be collected from different patients, and even from very different living systems (cell cultures, animals, patients). Each patient contributes to the description of the abstract system on only few types of observations. This is incomplete because there will be many gaps to bridge the different kinds of information related to functioning of the studied biological system.

To reach our objective, we explore the use of Semantic Web (SW) formalism which attracts a lot of interest in bioinformatics, to formalize knowledge and data. Data are observations of biological systems acquired within controlled experiments or in real life. Formalized knowledge is a representation of facts and rules acquired in a scientific domain, here medicine or life sciences. Applying machine learning techniques on data supports knowledge discovery, but it is only one particular source of knowledge. The methodological challenge is first to formalize the different types of available data within an abstract model of the biological system, and to integrate formalized knowledge in the model coming from medical literature and our medical expertise, including imaging or hospital data. By gathering a wide range of formalized data and knowledge within the same tool, we aim at creating a kind of abstract numerical twin that may be queried to infer new knowledge to assist drug design or drug repositioning.

On the longer run, the second challenge is to develop query answering at the abstract level but based on fragmented data. The objective is to answer queries about the numerical twin by exploiting the data coming from multiple patients. One of the difficulties is to detect groups of patients whose numerical twins are "similar to each other" (in a sense that remains to be defined). Semantic Web offers a natural framework for querying formalized data with multiple facets but may be limited by the time-efficiency of the query engines on a large number of patients. In such a context, numerical approaches (embedding) is more time-efficient but may lack accuracy. The challenge is to construct numerical representations in order to embed the data in a space in which the distances are both efficient to compute and semantically consistent with the applied notion of "similarity". Numerical machine learning techniques turn out to be an interesting perspective to address this challenge [54]. Recent research on advanced machine learning, such as representation learning, offers new perspectives to address our challenge. Our objective is to initiate collaborations with teams having strong backgrounds in machine learning (e.g. Ockham Inria Team) to propose innovative solutions. Another important point is the need for logic programming methodologies able to express complex queries, especially on heterogeneous or multimodal data. For neuroimaging, the availability of Neurolang for logic programming with heterogeneous data or NeuroQuery for query result consolidation based on automatic literature meta-analysis, for instance, should be very useful.

In most of the cases, the methodologies that we use to reach the above objectives are related to knowledge management/mining, formal reasoning, data mining or learning. Machine learning or deep learning approaches are probably less useful here. The main reason is related to the volume of available data. For rare disease like Niemann-Pick type C, for instance, the low prevalence means that 5 to 10 new patients are diagnosed in France each year, a number too low for deep neural networks. However, advances in transfer learning might be helpful here. For instance, a large number of brain pathologies come with dysfunction of intracellular cholesterol metabolism and storage. This is for instance the case of multiple sclerosis [53], for which large cohorts and databases are available worldwide. As a long term project, an interesting idea will be to leverage the large volume of data on multiple sclerosis to identify biomarkers of cholesterol dysfunction, e.g., in neuroimaging, and use transfer learning to adapt the network to Niemann-Pick type C patients.

# 5 New software, platforms, open data

## 5.1 New software

### 5.1.1 Chemfeat

**Name:** Chemfeat

**Keywords:** Python, Machine learning, Deep learning, Chemistry, Molecules

**Functional Description:** The project provides a Python package and command-line tool for generating feature vectors from molecules. The feature sets to include in the feature vectors are configurable via a simple YAML configuration file. Molecules are specified as lists of Inchis.

**URL:** https://gitlab.inria.fr/jrye/chemfeat

**Contact:** Jan-michael Rye

### 5.1.2 Hydronaut

**Keywords:** Machine learning, Deep learning, Python, Optimization, MLflow, Hydra, Optuna

**Functional Description:** A Python framework for machine- and deep-learning that makes it easy to use Hydra for hyperparameter configuration and MLflow for experiment tracking and result distribution. The user only needs to create a single YAML configuration file and a subclass of Hydronaut.Experiment to use the framework.

Hydra allows the user to systematically sweep all hyperparameter combinations or optimize them use different strategies with plugins for libraries such as Optuna.

MLflow provides a web interface, command-line interface and Python API for exploring and sharing the results.

The framework is fully compatible with PyTorch Lightning and provides a custom subclass to facilitate its use.

**URL:** https://gitlab.inria.fr/jrye/hydronaut

**Contact:** Jan-michael Rye

### 5.1.3 MLflow Extra

**Keywords:** MLflow, Python, Library

**Functional Description:** Utility scripts and matching python module for working with MLflow directories. It mainly provides functionality for moving around and regrouping "mlruns" directories which is useful in contexts such as migrating results from a cluster.

**URL:** https://gitlab.inria.fr/jrye/mlflow-extra

**Contact:** Jan-michael Rye

### 5.1.4 MolPred

**Keywords:** Hydronaut, ChemFeat, Chemistry, Molecules, Machine learning, Deep learning, Python

**Functional Description:** A Hydronaut-based framework that uses ChemFeat to generate feature vectors for training machine- and deep-learning models to predict properties of molecules.

**URL:** https://gitlab.inria.fr/jrye/molpred

**Contact:** Jan-michael Rye

### 5.1.5　SWoTTeD

**Name:** Sliding Window Tempral Tensor Decomposition

**Keywords:** Tensor decomposition, Temporal information, Machine learning

**Functional Description:** SWoTTeD is a tensor decomposition framework to extract temporal phenotypes from structured data. Most recent decomposition models allow extracting phenotypes that only describe snapshots of typical profiles, also called daily phenotypes. However, SWoTTeD extends the notion of daily phenotype into temporal phenotype describing an arrangement of features over a time window.

**URL:** https://hsebia.gitlabpages.inria.fr/swotted/readme.html

**Publication:** hal-03902300

**Contact:** Hana Sebia

**Participants:** Hana Sebia, Thomas Guyet

## 6　New results

### 6.1　New software tools for drug discovery

**Participants:**　Jan-Michael Rye.

We have developed two software tools to meet the internal needs of the team regarding in silico drug discovery. Both have been published online under open-source licences for the wider scientific community (MIT licence) and are available on Software Heritage. Both were presented at the seminar for digital health engineers at Inria Rennes in 2023.

- ChemFeat is an utility and extensible Python library for calculating chemical feature vectors for machine- and deep-learning models using external cheminformatics software packages.

- MolPred is a Hydronaut-base framework (see below) that integrates ChemFeat to create machine- and deep-learning models to predict the features of molecules.MolPred is a Hydronaut-based framework for building machine- and deep-learning predictors for molecular characteristics using Chemfeat.

### 6.2　New software tools for the management of machine- or deep-learning models.

**Participants:**　Jan-Michael Rye.

Two tools have been developed for the management of machine- or deep-learning models, including for the exploration of the hyperparameters. Here again, both have been published online under open-source licences for the wider scientific community (MIT licence) and are available on Software Heritage. Both were presented at the seminar for digital health engineers at Inria Rennes in 2022.

- Hydronaut Hydronaut is a framework for exploring the depths of hyperparameter space with Hydra and MLflow. Its goal is to encourage and facilitate the use of these tools while handling the sometimes unexpected complexity of using them together. Users benefit from both without having to worry about the implementation and are thus able to focus on developing their models.

- MLflow Extra is a set of utilities for managing MLflow data. It provides several commands for working with MLflow output directories, such as merging experiments from separate mlruns directories into a single one or fixing artifact paths after moving an mlruns directory.

## 6.3   SWoTTeD: An Extension of Tensor Decomposition to Temporal Phenotyping

**Participants:**    Hana Sebia, , Thomas Guyet.

Tensor decomposition has recently been gaining attention in the machine learning community for the analysis of individual traces, such as Electronic Health Records (EHR). However, this task becomes significantly more difficult when the data follows complex temporal patterns.

Our paper [10] introduces the notion of a temporal phenotype as an arrangement of features over time and it proposes SWoTTeD (Sliding Window for Temporal Tensor Decomposition), a novel method to discover hidden temporal patterns. SWoTTeD integrates several constraints and regularizations to enhance the interpretability of the extracted phenotypes. We validate our proposal using both synthetic and real-world datasets, and we present an original usecase using data from the Greater Paris University Hospital. The results show that SWoTTeD achieves at least as accurate reconstruction as recent state-of-the-art tensor decomposition models, and extracts temporal phenotypes that are meaningful for clinicians.

The implementation of SWoTTeD has been released (see Section 5). This work has been submitted to a journal (see pre-print arxiv.org/abs/2310.01201) and is now under revision.

## 6.4   Chronicles: Formalization of a Temporal Model

**Participants:**    Thomas Guyet.

This research line is intended as an introduction to the versatile model of chronicles for temporal data. Chronicles have been studied in the context of two analysis problems for temporal sequences: recognizing situations in temporal sequences and abstracting a set of temporal sequences. The first challenge benefits from the simple but expressive formalism to specify temporal behavior to match in a temporal sequence. The second challenge aims to abstract a collection of sequences by chronicles with the objective to extract characteristic behaviors.

Chronicles are closely related to temporal constraint networks. Not only do they share a similar graphical representation, they also have in common a notion of constraints in the timed succession of events. However, chronicles are definitely oriented towards fairly specific tasks in handling temporal data, by making explicit certain aspects of temporal data such as repetitions of an event.

We published a book [12] that first proposes a formal account of chronicles. Then, it exhibits an original lattice structure on the space of chronicles and proposes new counting approach for multiple occurrences of chronicle occurrences. This book also proposes a new approach for frequent temporal pattern mining using pattern structures. This latter proposal has been extended in [2] to address the problem of the interpretability of chronicle mining.

## 6.5   Robust Generation of Counterfactual Explanations

**Participants:**    Thomas Guyet.

Counterfactual explanations have become a mainstay of the explainable AI field. This particularly intuitive statement allows the user to understand what small but necessary changes would have to be made to a given situation in order to change a model prediction. The quality of a counterfactual depends on several criteria: realism, actionability, validity, robustness, etc. In the paper published in ECML 2023 [5], we are interested in the notion of robustness of a counterfactual.

More precisely, we focus on robustness to counterfactual input changes. This form of robustness is particularly challenging as it involves a trade-off between the robustness of the counterfactual and the

proximity with the example to explain. We propose a new framework, CROCO, that generates robust counterfactuals while managing effectively this trade-off, and guarantees the user a minimal robustness. An empirical evaluation on tabular datasets confirms the relevance and effectiveness of our approach.

In addition, [6, 9] presents an interactive visualization tool that exhibits counterfactual explanations to explain model decisions. Each individual sample is assessed to identify the set of changes needed to flip the output of the model. These explanations aim to provide end-users with personalized actionable insights with which to understand automated decisions. An interactive method is also provided so that users can explore various solutions. The functionality of the tool is demonstrated by its application to a customer retention dataset. The tool is compatible with any counterfactual explanation generator and decision model for tabular data.

## 6.6 Quantifying membrane binding and diffusion with fluorescence correlation spectroscopy diffusion laws

**Participants:**   Hugues Berry.

Many transient processes in cells arise from the binding of cytosolic proteins to membranes. Quantifying this membrane binding and its associated diffusion in the living cell is therefore of primary importance. Dynamic photonic microscopies, e.g., single/multiple particle tracking, fluorescence recovery after photobleaching, and fluorescence correlation spectroscopy (FCS), enable non-invasive measurement of molecular mobility in living cells and their plasma membranes. However, FCS with a single beam waist is of limited applicability with complex, non-Brownian, motions. Recently, the development of FCS diffusion laws methods has given access to the characterization of these complex motions, although none of them is applicable to the membrane binding case at the moment. In [3], we combined computer simulations and FCS experiments to propose an FCS diffusion law for membrane binding. First, we generated computer simulations of spot-variation FCS (svFCS) measurements for a membrane binding process combined to 2D and 3D diffusion at the membrane and in the bulk/cytosol, respectively. Then, using these simulations as a learning set, we derived an empirical diffusion law with three free parameters: the apparent binding constant KD, the diffusion coefficient on the membrane D2D, and the diffusion coefficient in the cytosol, D3D. Finally, we monitored, using svFCS, the dynamics of retroviral Gag proteins and associated mutants during their binding to supported lipid bilayers of different lipid composition or at plasma membranes of living cells, and we quantified KD and D2D in these conditions using our empirical diffusion law. Based on these experiments and numerical simulations, we conclude that this new approach enables correct estimation of membrane partitioning and membrane diffusion properties (KD and D2D) for peripheral membrane molecules.

**Participants:**   Hugues Berry, Audrey Denizot, Thomas Guyet, Benjamin Vidal, Luc Zimmer, Andrea Ducos, Florian Dupeuble, Hana Sebia, Mathieu Chambard, Guillaume Girard, Zoe Laffitte.

# 7   Bilateral contracts and grants with industry

## 7.1   Bilateral contracts with industry

AIstroSight is a joint project-team with the biotech company Theranexus. A plain "tutelle" of the team, Theranexus brings its research expertise in in vitro cell culture, disease modelling and imaging, both in terms of research workforce and data. The stand-alone "convention d'équipe-projet commune" of AIstroSight lists a group of 30 rare diseases of the central nervous system, that are of direct interest to Theranexus and thhat are associated with a specific regimen in terms of IP and legal affairs. However, AIstroSight members are allowed to work on pathologies outside this list without any restriction but with a different legal regimen vis-à-vis Theranexus.

# 8  Partnerships and cooperations

## 8.1  International initiatives

### 8.1.1  Participation in other International Programs
**ED-AIM**

> **Participants:**    Thomas Guyet.

**Title:** Ethical Design for Artificial Intelligence Models in patient management and treatment decisions

**Partner Institution(s):**    • CNRS, France

- Maison Française d'Oxford, UK

- Institute for the History of Representations and Ideas in Modernity (IHRIM), France

- IRISA, France

- Inria-Lyon, France

- Institute of Biomedical Engineering at Oxford University, UK

**Date/Duration:** 2021-2024

**Additionnal info/keywords:** The project, led by Thomas Guyet (INRIA), Mogens Lærke (MFO, IHRIM), and Pascal Marty (MFO) studies the deployment of Artificial Intelligence systems in healthcare settings from a philosophical perspective. This latter perspective is double, both epistemological and ethical. ED-AIM explores not only the ethical expectations of AI-based systems themselves but also the socio-technical context in which such systems are to be deployed. In a nutshell, the ambition is to address how to ethically design ethical AI-based systems in medicine with ontological, axiological, deontological and practical answers.

## 8.2  National initiatives
**ABC4M**

> **Participants:**    Nathan Quiblier, Hugues Berry.

**Title:** Approximate Bayesian computation-driven multimodal microscopy to explore the nuclear mobility of transcription factor

**Partner Institution(s):**    • Inria, Lyon (supervision)

- Institut Langevin, ESPCI, Paris

- Phlam laboratory, Lille

**Date/Duration:** 2020-2025

**Additionnal info/keywords:** Funded by the French National Agency for Research (ANR), Call "AAP2020" (grant ANR-20-CE45-0023-01). We combine computer simulations and Approximate Bayesian computation with simultaneous multiple microscopy methods (FCS and spt-PALM) to quantify the way transcription factors explore the nucleus to find their binding sites

**EngFlea**

> **Participants:**   Arnaud Hubert, Lisa Blum Moyse, Hugues Berry.

**Title:**  Engram of fast learning in the striatum

**Partner Institution(s):**   • CIRB, Collège de France, Paris (supervision)
  • Inria, Lyon (supervision)

**Date/Duration:**  2022-2026

**Additionnal info/keywords:**  Funded by the French National Agency for Research (ANR), Call "AAP2021" (grant ANR-21-CE16-0022-02). We study the link between endocannabinoid-mediated synaptic plasticity and fast learning of rodents thanks to a multidisciplinary approach combining in vitro and in vivo experimental neurophysiology with detailed subcellular biophysical models and large-scale neural network models.

**SecNet**

> **Participants:**   Schayma Ben Marzougui-El Garrai, Hugues Berry, Audrey Denizot.

**Title:**  Spatio-temporal dynamics of second messenger networks

**Partner Institution(s):**   • Institut de la Vision, Paris (supervision)
  • Inria, Lyon

**Date/Duration:**  2023-2026

**Additionnal info/keywords:**  Funded by the French National Agency for Research (ANR), Call "AAP2022" (grant 2023-ANR-22-CE16-0034-02). We combine cell biology approaches and mathematical modeling to provide a description of compartmentalized networks of second messengers that specifically regulate axon guidance and cell migration in response to repellent molecules.

**AIRACLES**

> **Participants:**   Thomas Guyet, Hana Sebia.

**Title:**  Chair AIRACLES (Fondation APHP)

**Partner Institution(s):**   • APHP, Paris
  • CentralSupelec, Saclay
  • Inria, Lyon

**Date/Duration:**  2020-2024

**Additionnal info/keywords:**  The AI-RACLES Chair in Artificial Intelligence, created in 2020 and co-directed by Etienne Audureau (AP-HP), Thomas Guyet (Inria), Laurent Le Brusquet and Arthur Tenenhaus (CentraleSupélec), aims at exploiting the massive data of the AP-HP's Health Data Warehouse (HDW) in order to carry out research focused on the exploration of the concept of vulnerability in health, whether related to ageing or to pathologies such as cancer or COVID-19.

**SAFEPaw**

> **Participants:**   Thomas Guyet.

**Title:** SAFEPaw (PEPR Santé Numérique)

**Partner Institution(s):**   • CNRS, Paris
- Université de Tours
- Ecole Normale Supérieure Paris-Saclay
- Aix-Marseille Université
- Ecole des Hautes Etudes En Santé Publique
- CHU Grenoble - Université Grenoble Alpes
- CHU Bordeaux - Université de Bordeaux
- Mines Saint-étienne
- Inria, Lyon

**Date/Duration:** 2023-2027

**Additionnal info/keywords:** The SAFEPaw project is a multidisciplinary project to question the improvement or optimization of care organization by distinguishing three points of view: Regulators / Patients / Healthcare professionals that includes doctors, health care institutions and ambulatory care. Our contribution to this project is to develop tools that would support decision making about the organization of care. It requires dually to be able to describe what is actually the current organization of care and to identify changes that may be improved or optimized. For that, we develop innovative visualization, data mining and operational research tools for care pathways analysis, management and planning. Their originality lays in their ability to consider three views of the care pathways: the patient, the regulator and the provider.

# 9 Dissemination

## 9.1 Promoting scientific activities

**Member of the organizing committees**

- Nathan Quiblier was member of the organizing commitee of YSN Imabio and of the steering committee of the thematic school Mifobio 2023, during which he co-organized a symposium called "Molecular dynamics: measurement and modeling".

- Hana Sebia participated in the organization of the GDR RSD summer school on distributed learning.

### 9.1.1 Scientific events: selection

**Chair of conference program committees**

- Thomas Guyet is co-chair of the Eigth AALTD workshop colocated with the conference ECML 2023.

### 9.1.2 Journal

**Member of the editorial boards**

- Hugues Berry is Associate Editor for *PLoS Computational Biology*.

- Thomas Guyet is member of the Editorial Committee of *Revue Ouverte d'Intelligence Artificielle*.

**Reviewer - reviewing activities**

- Thomas Guyet reviewed 1 article for the Journal of Information and Computation, and 2 articles for the journal of Machine Learning.

- Benjamin Vidal reviewed 1 article for the journal Neuroimage

- Audrey Denizot reviewed 1 article for the journal eLife

### 9.1.3 Invited talks

- Hugues Berry gave invited talks to the following seminars or workshops:

    - "Inferring Neural Networks from Electrophysiological and Functional Imaging", Centre de Recherches Mathématiques, November 2023, Montreal, Canada
    - "Journée 3R 2023", November 2023, Lyon
    - "Workshop IIT Dehli - Inria", October 2023, New Dehli, India
    - "Colloque 2023 du GDR ImaBio" July 2023, Paris
    - "6emes Journees Franco-Internationales d'Oncologie", June 2023, Paris
    - "Seminaires IA du Centre Leon Berard", May 2023, Lyon
    - "3rd 'Synaptic Microenvironment' Mini-symposium and Workshop", March 2023, Solden, Austria

- Audrey Denizot gave invited talks to the following conferences, seminars or workshops:

    - "Linking astrocyte nano-morphology to calcium activity at tripartite synapses", ICVS Satellite Symposium of the ISN-ESN Conference, August 2023, Braga, Portugal
    - "Tutorial - Modeling astrocytic calcium signaling", July 2023, Annual Computational Neuroscience Meeting, Leipzig, Germany
    - 'Dissecting the functions of astrocyte nano-architecture using voxel-based computational models", FENS Regional Meeting, May 2023, Algarve, Portugal
    - "Linking astrocyte morphology to calcium activity: insights from computational approaches" CEA Paris-Saclay, MIRCen seminar, April 2023, Fontenay-aux-Roses, France

- Thomas Guyet gave invited talks to the following seminars or workshops:

    - "Declarative Sequential Pattern Mining in ASP", November 2023, International Conference on Inductive Logic Programming, Bari, Italy.
    - "Modélisation et raisonnement sur les parcours de soins", École d'été interdisciplinaire en numérique de la santé, Juillet 2023, Sherbrooke, CA.
    - "Ethical and legal issues in the design and use of AI systems in health", eHealth and Ethics, Apr 2023, Nanterre, France.
    - "IA et données de santé : regards croisés sur les droits, devoirs et responsabilités des personnes publiques", Worshop Enjeux scientifiques et sociaux de l'IA, MITI/CNRS, Jan 2023, Paris, France

- Hana Sebia gave invited talks to the following seminar :

    - "Une extension de la décomposition tensoriel au phénotypage temporel", May 2023, Seminar of the Data Mining and Machine Learning research team in LIRIS, Lyon, France

- Benjamin Vidal gave an invited talk to the following seminar :

    - "Une plateforme expérimentale préclinique in vivo pour les médicaments en neurologie", November 2023, Journée Scientifique de la Structure Fédérative de Recherche Santé Lyon-Est, France

### 9.1.4  Poster presentation

Audrey Denizot presented posters during the following international conferences:

- "Linking astrocyte function to cell shape: insights from computational models", 50th Naito conference, "Glia World - Glial Cells Governing Brain Functions", October 2023, Sapporo, Japan

- "Computational tools to unravel mechanistic links between intracellular architecture and cell function", XVI European Meeting on Glial Cells in Health and Disease, GLIA 2023, July 2023, Berlin, Germany

### 9.1.5  Leadership within the scientific community

- Hugues Berry has been Deputy Scientific Director of Inria, until September 2023

- Hugues Berry has served in the HCERES visiting committee of CAPS Lab, U1093 INSERM, Dijon and Physico-ChimieCurie, UMR CNRS 168, Institut Curie, Paris

- Hugues Berry is a member of Inserm's scientific selection committee for "Technologies de la santé" (CSS7)

- Thomas Guyet is board member of the French Association for Artificial Intelligence, chair of the steering commity of the French Platform of Artificial Intelligence.

- Thomas Guyet is member of the steering committee of the TIME Symposium

### 9.1.6  Scientific expertise

- Hugues Berry is a member of the Comité d'Expertise et Scientifique pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (CESREES)

- Hugues Berry has been a member of the Inria Lyon CRCN selection committee and the Inserm CSS7 DR2 selection committee

- Thomas Guyet reviewed projects for the BRGM and the Université Grenoble Alpes (IRGA call)

- Thomas Guyet is member of the national evaluation committee of Inrae Engineers (section Computer Science)

- Audrey Denizot is a member of the ANR Comité d'Evaluation scientifique 45 : "Interfaces : mathématiques, sciences du numérique - biologie, santé"

## 9.2  Teaching - Supervision - Juries

- Thomas Guyet was teaching Data Mining in the Master "Informatique Fondamentale" (IFA/ENS Lyon)

- Hugues Berry and Thomas Guyet contributed to the course "AI in Health" of the master "Recherche Biomédicale" of the Lyon's medicine faculty.

- Nathan Quiblier was teaching fellow in game theory and information in the Master "Économétrie et statistiques" (ISFA/Université Claude Bernard-Lyon 1)

- Nathan Quiblier was teaching fellow in insurance's economy in the Master "Actuariat" (ISFA/Université Claude Bernard-Lyon 1)

- Hana Sebia was a teaching fellow in the course 'Deep Learning' of the Master "Data Science" of Claude Bernard Lyon 1 University

### 9.2.1   Supervision

PhD. Students

- **Andréa Ducos**, "Partial differential equation discovery for spatio-temporal simulations in cells", since 02/11/2023, supervised by T. Guyet, A. Denizot and H. Berry

- **Schayma Ben Marzougui**, "Modeling the spatio-temporal dynamics of second messenger networks", since 01/10/2023, supervised by A. Denizot and H. Berry

- **Florian Dupeuble**, "Biophysical modeling of neurovascular coupling at the gliovascular unit", since 01/09/2023, supervised by A. Denizot and H. Berry

- **Victor Guyomard**, "Explanation of decisions taken by machine learning algorithms", defended on 23/11/2023, supervised by T. Bouadi, F. Fessant, T. Guyet

- **Arnaud Hubert**, " Modelling endocannabinoid-mediated synaptic plasticity and its implication in fast learning", since 01/11/2022, supervised by H. Berry

- **Eric Pardoux**, "Ethical issues in the use of artificial intelligence in healthcare: the contribution of an epistemological perspective" since 01/10/2021, supervised by T. Guyet and M. Laerke (CNRS, MFO)

- **Thibaut Peyric**, "Single-cell multi-omics data integration for gene regulatory network inference", since 01/11/2023, supervised by A. Crombach (Inria/Beagle) and T. Guyet.

- **Nathan Quiblier**, "Intracellular signaling: multi-scale modeling and statistical learning methods", since 01/11/2021, supervised by H. Berry

- **Hana Sebia**, "Deep phenotyping of patients" since 01/11/2022, supervised by T. Guyet and H. Berry

Master Students

- **Mathieu Chambard**, "Multiscale Modeling with Partial Differential Equations", from 02/2023 to 07/2023, supervised by T. Guyet, A. Denizot and H. Berry

- **Zoë Laffitte**, "Implementation of a database of 3D cell reconstructions for reaction-diffusion simulations", from 04/2023 to 07/2023 and 11/2023-07/2024, supervised by A. Denizot and J. M. Rye

- **Guillaume Girard**, " ", from 04/2023 to 07/2023, supervised by N. Simon

### 9.2.2   Juries

- Hugues Berry has been a member of the PhD committees of M. Davy (Univ. Montpellier, France, Chair) and A. Poshtokhi (Univ. of Ulster, UK, external reviewer)

- Thomas Guyet has been a member of the PhD juries of J. Richard (Univ. La Rochelle, France, reviewer) and J. Loisel (Univ. of Paris-Saclay, reviewer)

- Thomas Guyet is member of the PhD Advisory Committee of A. Khudiyev (Univ. Strasbourg), I. Azizi (Univ. Paris Cité), J. Aalmoes (Inria Lyon), M. Bhan (LIP6), N. Mountasir (Univ. Strasbourg) and Y. Oubelmouh (Univ. Tours).

- Audrey Denizot has been a member of the 1st year PhD evaluation committee of Den-Whilrex Garcia, Paris-Saclay and of the 2nd year PhD evaluation committee of Aitakin Ezzati, Université Aix-Marseille.

## 9.3   Popularization

### 9.3.1   Education

- Audrey Denizot is a member of the board of directors and of the editorial committee of the association "Papier Maché Sciences".

- Benjamin Vidal produced online resources about ultrasound imaging for the "Culture Sciences Physique" website (to be published soon)

### 9.3.2   Interventions

- Andrea Ducos took part in the organization of a day "Filles et informatique : une équation lumineuse !" at ENS Lyon. The goal was to inform a hundred high-school girls about careers of women related to computing and mathematics, so that they can consider these paths more favorably when choosing their career path

- Arnaud Hubert took part to a workshop "Pourquoi faire une thèse en informatique ?" at the INSA's library Marie Curie during the "Festival numérique" organised by the Inria Lyon center. The goal was to inform and exchange with students about many topics related to Phd.

# 10   Scientific production

## 10.1   Publications of the year

**International journals**

[1]   C. Gautrais, P. Cellier, T. Guyet, R. Quiniou and A. Termier. 'Sky-signatures: detecting and characterizing recurrent behavior in sequential data'. In: *Data Mining and Knowledge Discovery* (29th Aug. 2023). DOI: 10.1007/s10618-023-00949-1. URL: https://hal.science/hal-04401641.

[2]   T. Guyet. 'Admissible generalisation of temporal sequences as chronicles'. In: *Journal of Applied Non-Classical Logics* 33.3-4 (7th Aug. 2023), pp. 641–653. DOI: 10.1080/11663081.2023.2244717. URL: https://hal.science/hal-04383648.

[3]   A. Mouttou, E. Bremaud, J. Noero, R. Dibsy, C. Arone, J. Mak, D. Muriaux, H. Berry and C. Favard. 'Quantifying membrane binding and diffusion with fluorescence correlation spectroscopy diffusion laws'. In: *Biophysical Journal* 122.11 (June 2023), pp. 2216–2229. DOI: 10.1016/j.bpj.2023.01.006. URL: https://inria.hal.science/hal-04165544.

**Invited conferences**

[4]   T. Guyet. 'Declarative Sequential Pattern Mining in ASP'. In: Conference on Inductive Logic Programming (ILP). Bari, Italy, 13th Nov. 2023. URL: https://inria.hal.science/hal-04295035.

**International peer-reviewed conferences**

[5]   V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Generating robust counterfactual explanations'. In: ECML/PKDD - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Turin (Italie), Italy, 2023, pp. 1–16. URL: https://hal.science/hal-04255500.

[6]   V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Interactive Visualization of Counterfactual Explanations for Tabular Data'. In: ECML/PKDD 2023 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vol. 14175. Lecture Notes in Computer Science. Turin, Italy: Springer Nature Switzerland, 17th Sept. 2023, pp. 330–334. DOI: 10.1007/978-3-031-43430-3_25. URL: https://hal.science/hal-04255496.

[7]    P. N. Sawadogo, T. Guyet and E. Audureau. 'Interface d'interrogation graphique de parcours de soins à base de chroniques'. In: *Revue des Nouvelles Technologies de l'Information*. Conférence Extraction et gestion de connaissances. RNTI-E-39. Lyon, France, 17th Jan. 2023. URL: `https://hal.science/hal-04181253`.

### National peer-reviewed Conferences

[8]    T. Guyet. 'Analyse d'une enquête sur la sémantique des motifs séquentiels avec négation'. In: *Actes CNIA 2023 - Conférence Nationale en Intelligence Artificielle, PFIA*. CNIA 2023 - Conférence Nationale en Intelligence Artificielle, PFIA. Strasbourg, France, 3rd July 2023, pp. 62–71. URL: `https://hal.science/hal-04164278`.

[9]    V. Guyomard, F. Wallyn, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Visualiser des explications contrefactuelles pour des données tabulaires'. In: *Revue des Nouvelles Technologies de l'Information*. EGC 2023 - Conférence Extraction et Gestion des Connaissances. Vol. RNTI-E-39. Lyon, France, 16th Jan. 2023, pp. 557–564. URL: `https://hal.science/hal-04181257`.

[10]   H. Sebia, T. Guyet and E. Audureau. 'Une extension de la décomposition tensorielle au phénotypage temporel'. In: EGC 2023 - 23ème Conférence Francophone sur l'Extraction et Gestion des Connaissances. Lyon, France, 16th Jan. 2023. URL: `https://inria.hal.science/hal-03902300`.

### Scientific books

[11]   E. Adam, T. Guyet, B. Le Blanc, D. Longin, Z. Bouraoui, S. Bringay, R. Gaudel, C. Laclau, C. Launois, M. Morge, C. Roussey, F. Schwarzentruber, C. Trojahn, É. Vareilles and A. Wilczynski. *Conférence Nationale d'Intelligence Artificielle (CNIA) Année 2023 : Activité AFIA 1er août 2021 – 31 juillet 2023*. Association Française pour l'Intelligence Artificielle, 4th Oct. 2023, pp. XLVI, 1–123. URL: `https://ut3-toulouseinp.hal.science/hal-04310171`.

[12]   T. Guyet and P. Besnard. *Chronicles: Formalization of a Temporal Model*. SpringerBriefs in Computer Science (BRIEFSCOMPUTER). Springer International Publishing, 26th July 2023, p. 121. DOI: `10.1007/978-3-031-33693-5`. URL: `https://inria.hal.science/hal-04173959`.

### Scientific book chapters

[13]   H. Berry. 'Les utilisations de l'IA pour l'hôpital et la santé publique : données nécessaires, applications et obstacles potentiels'. In: *Maîtriser l'IA au service de l'action publique*. Berger-Levrault, Jan. 2023. URL: `https://inria.hal.science/hal-04165489`.

### Edition (books, proceedings, special issue of a journal)

[14]   G. Ifrim, R. Tavenard, A. Bagnall, P. Schaefer, S. Malinowski, T. Guyet and V. Lemaire, eds. *Advanced Analytics and Learning on Temporal Data*. AALTD 2023 - 8th Workshop on Advanced Analytics and Learning on Temporal Data. Vol. 14343. Lecture Notes in Computer Science. Springer Nature Switzerland, 2023. DOI: `10.1007/978-3-031-49896-1`. URL: `https://inria.hal.science/hal-04383684`.

### Reports & preprints

[15]   C. Alamichel, J. Calvo, E. Hingant, S. Latrach, N. Quiblier and R. Yvinec. *Modeling compartmentalization within intracellular signaling pathway*. 12th Jan. 2024. URL: `https://hal.science/hal-04098543`.

[16]   S. Sadoune, A. Fraboulet, A. Richard, F. Talbot, L. Boussel and H. Berry. *Analyse automatique de négations pour la radiologie et autres textes cliniques en français par modèles de langage*. 18th July 2023. URL: `https://inria.hal.science/hal-04165319`.

**Other scientific publications**

[17]  T. Guyet. *Ethical and legal issues in the design and use of AI systems in health.* Nanterre, France, 13th Apr. 2023. URL: https://inria.hal.science/hal-04068428.

[18]  H. Sebia. *SWoTTeD : An Extension of Tensor Decomposition to Temporal Phenotyping.* 27th Nov. 2023. URL: https://hal.science/hal-04310487.

## 10.2   Other

**Softwares**

[19]  [SW] J. M. Rye, *ChemFeat* version 2023.11, 18th Oct. 2023. Inria Lyon. LIC: http://spdx.org/licenses/MIT. HAL: ⟨hal-04248197⟩, URL: https://hal.science/hal-04248197, VCS: https://gitlab.inria.fr/jrye/chemfeat, SWHID: ⟨swh:1:dir:97647e8cc377bcb350f3169bdf1218da6f0152bf;origin=https://gitlab.inria.fr/jrye/chemfeat.git;visit=swh:1:snp:02443124ed4ee0d8d724fefd38bf9b271361cc09;anchor=swh:1:rev:25bfa1d9d55358292cbdf185f96f20c4d9f17893⟩.

[20]  [SW] J.-M. Rye, *Hydronaut* version 2023.17, 17th Oct. 2023. Inria Lyon. LIC: http://spdx.org/licenses/MIT. HAL: ⟨hal-04246627⟩, URL: https://hal.science/hal-04246627, VCS: https://gitlab.inria.fr/jrye/hydronaut, SWHID: ⟨swh:1:dir:9e229160ec54cad819527ce8726ddff662933a5c;origin=https://gitlab.inria.fr/jrye/hydronaut.git;visit=swh:1:snp:bf5f6a804696373ea8ad4b34e7d25a93ea922185;anchor=swh:1:rev:39fe2327cfc98544d039d53c0189fc4469dfac3f⟩.

[21]  [SW] J.-M. Rye, *MolPred* version 2023.5, 18th Oct. 2023. Inria Lyon. LIC: http://spdx.org/licenses/MIT. HAL: ⟨hal-04248217⟩, URL: https://hal.science/hal-04248217, VCS: https://gitlab.inria.fr/jrye/molpred, SWHID: ⟨swh:1:dir:30804037d1dfde5004422d725625b765d5e68101;origin=https://gitlab.inria.fr/jrye/molpred.git;visit=swh:1:snp:401a0b556180686321959104732de8127df4f235;anchor=swh:1:rev:5dd0d37ab787a10a0b3baa55e685c7c83a8071f0⟩.

## 10.3   Cited publications

[22]  S. Azam, M. Haque, M. Jakaria, S.-H. Jo, I.-S. Kim and D.-K. Choi. 'G-Protein-Coupled Receptors in CNS: A Potential Therapeutic Target for Intervention in Neurodegenerative Disorders and Associated Cognitive Deficits'. In: *Cells* 9.2 (2020), p. 506.

[23]  A. Badoual, M. Arizono, A. Denizot, M. Ducros, H. Berry, U. Valentin Nägerl and C. Kervrann. 'Simulation of Astrocytic Calcium Dynamics in Lattice Light Sheet Microscopy Images'. In: *IEEE International Symposium on Biomedical Imaging, ISBI.* 2021, pp. 135–139.

[24]  N. Bazargani and A. Attwel. 'Astrocyte calcium signaling: the third wave'. In: *Nature Neuroscience* 19 (2016), pp. 182–189.

[25]  A. Bespalov, T. Steckler, B. Altevogt, E. Koustova, P. Skolnick, D. Deaver, M. Millan, J. Bastlund, D. Doller, J. Witkin, P. Moser, P. O'Donnell, U. Ebert, M. Geyer, E. Prinssen, T. Ballard and M. Macleod. 'Failed trials for central nervous system disorders do not necessarily invalidate preclinical models and drug targets'. In: *Nature Reviews Drug Discovery* 15 (2016), p. 516.

[26]  T. Bie. 'Subjective interestingness in exploratory data mining'. In: *International Symposium on Intelligent Data Analysis.* 2013, pp. 19–31.

[27]  M. Bienvenu. 'Ontology-mediated query answering: harnessing knowledge to get more from data'. en. In: *Proceedings IJCAI'16.* 2016, pp. 4058–4061.

[28]  S. Brunton, J. Proctor and J. Kutz. 'Discovering governing equations from data by sparse identification of nonlinear dynamical systems'. In: *Proc Natl Acad Sci USA* 113 (2016), p. 3932.

[29]  A. Callahan. 'ACE: the Advanced Cohort Engine for searching longitudinal patient records'. In: *J Am Med Inform Assoc* 28.7 (2021), pp. 1468–1479.

[30] E. Chin and E. Goh. 'MeCP2 Dysfunction in Rett Syndrome and Neuropsychiatric Disorders'. In: *Psychiatric Disorders: Methods and Protocols, Methods in Molecular Biology*. Ed. by F. Kobeissy. Vol. 2011. Humana, New York, NY, 2019, pp. 573–592.

[31] A. Cougnoux, J. Yerger, M. Fellmeth, J. Serra Vinardell, K. Martin, F. Navid, J. Iben, C. Wassif, N. Cawley and F. Porter. 'Single Cell Transcriptome Analysis of Niemann-Pick Disease, Type C1 Cerebella'. In: *Int J Mol Sci* 21 (2020), p. 5368.

[32] Y. Cui, V. Paille, H. Xu, S. Genet, B. Delord, E. Fino, H. Berry and L. Venance. 'Endocannabinoids mediate bidirectional striatal spike-timing dependent plasticity'. In: *J Physiol* 593.13 (2015), pp. 2833–2849.

[33] Y. Cui, I. Prokin, A. Mendes, H. Berry and L. Venance. 'Robustness of STDP to spike timing jitter'. In: *Scientific Reports* 8 (2018), p. 8139.

[34] Y. Cui, I. Prokin, H. Xu, B. Delord, S. Genet, L. Venance and H. Berry. 'Endocannabinoid dynamics gate spike-timing dependent depression and potentiation'. In: *eLife* 5 (2016), e13185.

[35] Y. Cui, Y. Yang, Z. Ni, Y. Dong, G.-H. Cai, A. Foncelle, S. Ma, K. Sang, S. Tang, Y. Li, Y. Shen, H. Berry, S.-X. Wu and H. Hu. 'Astroglial-Kir4.1 in lateral habenula drives neuronal bursts in depression'. In: *Nature* 554 (2018), pp. 323–327.

[36] A. Denizot, M. Arizono, V. U. Nagerl, H. Soula and H. Berry. 'Simulation of calcium signaling in fine astrocytic processes: effect of spatial properties on spontaneous activity'. In: *PLoS Comput Biol* 15.8 (2019), p. 1006795.

[37] A. Denizot, C. Cali, H. Berry and E. De Schutter. 'Stochastic spatially-extended simulations predict the effect of ER distribution on astrocytic microdomain Ca2+ activity'. In: *ACM NanoCom* 20 (2021), pp. 1–5.

[38] S. Erikainen and S. Chan. 'Contested futures: envisioning "Personalized', Stratified," and "Precision" medicine'. In: *New Genet Soc* 38.3 (2019), pp. 308–330.

[39] T. Filzen, P. Kutchukian, J. Hermes, J. Li and M. Tudor. 'Representing high throughput expression profiles via perturbation barcodes reveals compound targets'. In: *PLoS Comput Biol* 13.2 (2017), p. 1005335.

[40] S. Finkbeiner. 'Bridging the Valley of Death of therapeutics for neurodegeneration'. In: *Nature Medicine* 16 (2010), pp. 1227–1232.

[41] A. Foncelle, A. Mendes, J. Jedrzejewska-Szmek, S. Valtcheva, H. Berry, K. Blackwell and L. Venance. 'Modulation of spike-timing dependent plasticity: towards the inclusion of a third factor in computational models'. In: *Frontiers in Computational Neuroscience* 12 (2018), p. 49.

[42] J. Freitas, K. Johnson, E. Golden, G. Nadkarni, E. Bottinger, B. Glicksberg and R. Miotto. 'Phe2vec: Automated disease phenotyping based on unsupervised embeddings from electronic health records'. In: *Patterns* 2.9 (2021), p. 100337.

[43] K. Friston, A. Mechelli, R. Turner and C. Price. 'Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics'. In: *NeuroImage* 12 (2000), pp. 466–477.

[44] N. J. Gamo, M. R. Birknow, D. Sullivan, M. A. Kondo, Y. Horiuchi, T. Sakurai, B. S. Slusher and A. Sawa. 'Valley of death: A proposal to build a "translational bridge" for the next generation'. In: *Neurosci Res* 115 (2017), pp. 1–4.

[45] G. Gangarossa, S. Perez, Y. Dembitskaya, I. Prokin, H. Berry and L. Venance. 'BDNF controls bidirectional endocannabinoid-plasticity at corticostriatal synapses'. In: *Cerebral Cortex* 1.30 (2020), pp. 197–217.

[46] R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. Ambasta and P. Kumar. 'Artificial intelligence to deep learning: machine intelligence approach for drug discovery'. In: *Molecular Diversity* 25 (2021), pp. 1315–1360.

[47] S. Hodavance, C. Gareri, R. Torok and H. Rockman. 'G Protein-coupled Receptor Biased Agonism'. In: *J Cardiovasc Pharmacol* 67.3 (2016), pp. 193–202.

[48] J. Ip, N. Mellios and M. Sur. 'Rett syndrome: insights into genetic, molecular and circuit mechanisms'. In: *Nat Rev Neurosci* 19 (2018), pp. 368–382.

[49] R. Jolivet, J. Coggan, I. Allaman and P. Magistretti. 'Multi-timescale Modeling of Activity-Dependent Metabolic Coupling in the Neuron-Glia-Vasculature Ensemble'. In: *PLoS Comput Biol* 11 (2015), p. 1004036.

[50] M. Kiernan, S. Vucic, K. Talbot, C. McDermott, O. Hardiman, J. Shefner, A. Al-Chalabi, W. Huynh, M. Cudkowicz, P. Talman, L. Berg, T. Dharmadasa, P. Wicks, C. Reilly and M. Turner. 'Improving clinical trial outcomes in amyotrophic lateral sclerosis'. In: *Nature Reviews Neurology* 17 (Dec. 2020).

[51] C. Kloft, M. Trame and C. Ritter. 'Systems pharmacology in drug development and therapeutic use — A forthcoming paradigm shift'. In: *Eur J Pharm Sci* 94 (2016), pp. 4–14.

[52] N. Lim and P. Pavlidis. 'Evaluation of connectivity map shows limited reproducibility in drug repositioning'. In: *Sci Rep* 11.1 (2013), p. 17624.

[53] B. Lőrincz, E. Jury, M. Vrablik, M. Ramanathan and T. Uher. 'The role of cholesterol metabolism in multiple sclerosis: From molecular pathophysiology to radiological and clinical disease activity'. In: *Autoimmunity Rev* 21.6 (2022), p. 103088.

[54] P. Monnin, C. Raïssi, A. Napoli and A. Coulet. 'Discovering alignment relations with Graph Convolutional Networks: a biomedical case study'. In: 2021, p. 03452182.

[55] A. Newman-Tancredi. 'Translating biased agonists from molecules to medications: Serotonin 5-HT1A receptor functional selectivity for CNS disorders'. In: *Pharmacology Therapeutics* 107937 (2021).

[56] P. O'Donnell, L. B. Rosen, R. Alexander, V. N. Murthy, C. H. Davies and E. Ratti. 'Strategies to Address Challenges in Neuroscience Drug Discovery and Development'. In: *International Journal of Neuropsychopharmacology* 22 (2019), pp. 445–448.

[57] V. Pérez-Nueno. 'Using quantitative systems pharmacology for novel drug discovery'. In: *Expert Opin Drug Discov* 10.12 (2015), pp. 1315–1331.

[58] S. Pfeffer. 'NPC intracellular cholesterol transporter 1 (NPC1)-mediated cholesterol export from lysosomes'. In: *J Biol Chem* 294.5 (2019), pp. 1706–1709.

[59] Z. Qian, K. Kacprzyk and M. Schaar. 'D-code: discovering closed-form ODEs from observed trajectories'. In: *Proceedings ICLR 2021*. 2021.

[60] M. Raissi, P. Perdikaris and G. Karniadakis. 'Physics-informed neural networks : A deep-learning framework for solving forward and inverse problem involving nonlinear partial differential equations'. In: *J Comput Phys* 378 (2019), pp. 686–707.

[61] C. Shivade, P. Raghavan, E. Fosler-Lussier, P. Embi, N. Elhadad, S. Johnson and A. Lai. 'A review of approaches to identifying patient phenotype cohorts using electronic health records'. In: *J Am Med Inform Assoc* 21.2 (2014), pp. 221–230.

[62] S. Shovlin and D. Tropea. 'Transcriptome level analysis in Rett syndrome using human samples from different tissues'. In: *Orphanet J Rare Dis* 13.1 (2018), p. 113.

[63] A. Silberschatz and A. Tuzhilin. 'On subjective measures of interestingness in knowledge discovery'. In: *Proceedings KDD'95*. 1995, pp. 275–281.

[64] T. Takeuchi, A. Duszkiewicz and R. Morris. 'The synaptic plasticity and memory hypothesis: encoding, storage and persistence'. In: *Philos Trans R Soc Lond B* 369 (2014), p. 20130288.

[65] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer and S. Zhao. 'Applications of machine learning in drug discovery and development'. In: *Nature Reviews Drug Discovery* 18 (2019), pp. 463–477.

[66] M. Vanier. 'Niemann-Pick disease type C'. In: *Orphanet J Rare Dis* 5.16 (2010).

[67] A. Verkhratsky and M. Nedergaard. 'Physiology of Astroglia'. In: *Physiol Rev* 98.1 (2018), pp. 239–289.

[68]    B. Vidal, S. Fieux, J. Redouté, M. Villien, F. Bonnefoi, D. Bars, A. Newman-Tancredi, N. Costes and L. Zimmer. 'In vivo biased agonism at 5-HT1A receptors: characterisation by simultaneous PET/MR imaging'. In: *Neuropsychopharmacology* (2018), pp. 2310–2319.

[69]    R. Weinshilboum. 'Inheritance and Drug Response'. In: *N. Engl J Med* 348 (2003), pp. 529–537.

[70]    H. Xu, S. Perez, A. Cornil, B. Detraux, I. Prokin, Y. Cui, B. Degos, H. Berry, A. De Kerchove D'exaerde and L. Venance. 'Dopamine- endocannabinoid interactions mediate spike-timing dependent potentiation in the striatum'. In: *Nature Communications* 9 (2018), p. 4118.

[71]    K. Yin, D. Qian, K.-W. Cheung, B. Fung and J. Poon. 'Learning phenotypes and dynamic patient representations via rnn regularized collective non-negative tensor factorization'. In: *Proceedings of the 2019 AAAI Conference on Artificial Intelligence*. 2019, pp. 1246–1253.

[72]    M. Zhang, D. Strnatka, C. Donohue, J. Hallows, I. Vincent and R. Erickson. 'Astrocyte-only Npc1 reduces neuronal cholesterol and triples life span of Npc1-/- mice'. In: *J Neurosci Res* 86 (2008), pp. 2848–2856.